Multivariate analysis and neural networks application to price forecasting in the Brazilian agricultural market

Carlos Alberto Orge Pinheiro[1] Valter de Senna[2]

[1]Universidade do Estado da Bahia (UNEB), 411500-000, Salvador, BA, Brasil. E-mail: email@carlospinheiro.com.br. Corresponding author.
[2]Faculdade de Tecnologia Senai Cimatec, Salvador, BA, Brasil.

ABSTRACT: *The purpose of this study is to apply the methodology proposed by PINHEIRO & SENNA (2015) to a set of agricultural products traded in Brazil. The multivariate and nonlinear character of this methodology has shown to be suitable, as compared to the neural network model, since it allows for a better predictive performance. Results obtained in an out-of-sample period, by using the calculated error and statistical test, confirmed this statement. This study will be useful to farmers as price forecasting based on their tendency is relevant.*
Key words: *neural networks, multivariate analysis, agricultural products, forecasting.*

Aplicação de análise multivariada e redes neurais para previsão de preços no mercado agrícola brasileiro

RESUMO: *O objetivo deste estudo é aplicar a metodologia proposta por PINHEIRO & SENNA (2015) a um conjunto de produtos agrícolas negociados no Brasil. O caráter multivariado e não linear dessa metodologia mostrou ser adequado, quando comparado com o modelo de redes neurais, pois ele permite um melhor desempenho preditivo. Os resultados obtidos em um período fora da amostra, usando o erro calculado e o teste estatístico, confirmam isso. Este estudo será útil aos agricultores, pois a previsão dos preços com base em sua tendência é relevante.*
Palavras-chave: *redes neurais, análise multivariada, produtos agrícolas, previsão.*

INTRODUCTION

In recent years, agriculture in Brazil has developed and modernized, especially as an activity of high economic and social value, with a marked tendency to grow (CAMPOS, 2007). Internal and external development of the markets contributed to the dynamic character of agriculture, which incorporated productive technologies to meet the requirements of these markets.

For farmers, price analysis has a singular importance as a component of the exchange mechanism. Therefore, decision making requires that price behavior is known before the harvest is done (RIBEIRO et al., 2010).

The main problem in forecasting agricultural prices is in the seasonality that occurs due to the weather, market, and conjunctural factors, which cause income uncertainty for farmers (MARTINS & MARTINELLI, 2010). Thus, farmers and academic people have been increasingly interested in price forecasting as it allows that price trend uncertainty is reduced.

In the literature (YONENAGA & FIGUEREDO, 1998; BRESSAN, 2004; SOBREIRO et al., 2008; LIMA et al., 2010; RIBEIRO et al., 2010; CERETTA et al., 2010; FERREIRA et al., 2011), the artificial neural networks (ANN) and classic (autoregressive) models are applied in the context of the Brazilian market to identify seasonal patterns and predict prices of agricultural products. In addition to these applications for the international market of agricultural commodities, some authors (SHAHWAN & ODENING, 2007; LI et al., 2010; JHA & SINHA, 2013), use the ANN model as an alternative to forecast agricultural prices.

However, a gap exists in these studies as they neglect the set of prices of agricultural products to favor structure capturing in the study time series (HASSANI & MAHMOUDVAND, 2013).

So, the purpose of this study was to apply the methodology proposed by PINHEIRO & SENNA (2015), which combines the ANN model in forecasting prices of agricultural products with the multivariate singular spectrum analysis (MSSA)

model. The latter model captures the structure of the study time series, based on the set of other time series; thus, supporting the agricultural sector. Therefore, the methodology efficiency was evaluated by means of prediction performance.

MATERIALS AND METHODS

This study on time series of prices for agricultural products was performed based on the commodities identified bellow as SUG (sugar), COT (cotton) COR (corn), COF (coffee), and SOY (soy). These products were chosen based on the export volume growth observed in the last years, according to data available at the Brazilian Ministry of Agriculture (MAPA, 2014). Daily prices were obtained from the database of the Center for Advanced Studies in Applied Economics (CEPEA, for *Centro de Estudos Avançados em Economia Aplicada*) and correspond to the period from Jan 13, 2012 to Dec 20, 2013. Then, the time series was converted to weekly periodicity, totaling 96 weeks.

The methodology is the same as that proposed by PINHEIRO & SENNA (2015), which will be named from now on as ANN-MSSA. In addition to it, the ANN model was also applied for comparison between predictive performances as in the studies by YONENAGA & FIGUEREDO (1998); BRESSAN (2004); SHAHWAN & ODENING (2007); SOBREIRO et al. (2008); LIMA et al. (2010); RIBEIRO et al. (2010); CERETTA et al. (2010); LI et al. (2010); FERREIRA et al. (2011); JHA & SINHA (2013).

Artificial neural network (ANN) model

The ANN model is a nonparametric forecasting model, which does not require normality, stationarity, and linearity of data assumptions. In a simplified manner, it is a computational structure based on a biological process inspired by the human brain architecture. In situations in which the law of data generation is unknown, the ANN model becomes useful because it can approach to any non-linear function. Evidence on the nonlinearity of time series of agricultural products in the Brazilian market can be seen in the studies by LIMA et al. (2010); CERETTA et al. (2010); FERREIRA et al. (2011) e BECKMANN & CZUDAJ (2014) for the international market. This evidence justifies the use of the ANN model in the present study.

Elements of the artificial neuron are represented by: m that indicates the number of the neuron input signals; x_j the j-th neuron input signal;

w_{gj} the weight associated with the j-th neuron input signal g; b the threshold of each neuron, also called bias; v_g a weighted combination of input signals and the g-th neuron bias and $\varphi(.)$ as an activation function of the g-th neuron (HAYKIN, 2001).

Replacing the bias b_g for a fixed input $x_0=1$ is possible, so that the bias becomes a new synaptic weight $W_{g0}=b_g$. Thus, the neuron g is described as (PASQUOTTO, 2010):

$$v_g = \sum_{j=0}^{m} x_j w_{gj}$$

(1)

and

$$y_g = \varphi(v_g(t))$$

(2)

with v_g defined as an induced local field or activation potential, and the activation function defining the output $y_g(t)$ of the g-th neuron at instant t.

Classification given in the literature for the way neurons are distributed in the network considered that they are arranged in layers. According to HAYKIN (2001), the neural network architecture can be disposed in monolayer, multilayer, fed forward, and recurrent. Proper use of the ANN model requires training to interactively adjust the network parameters to a sequence of events.

The supervised training works indicated the correct answer for each situation at the network output. For this purpose, a set of input data is presented to the neural network as an example. It generates a network output, which is compared with the expected output, thus giving the corresponding error.

Multivariate singular spectrum (MSS) model

As the ANN model, the MSS model is also non-parametric. In its presentation, it consisted of two complementary stages: decomposition and reconstruction. Decomposition stage is defined by two steps: incorporation and decomposition in singular value. Incorporation is a mapping that transfers a set M of one-dimension time series $Y_{N_i}^{(i)} = y_1^{(i)},...,y_{N_i}^{(i)}$ to a multidimensional array $\left[X_1^{(i)},...,X_{K_i}^{(i)}\right]$ with lagged vectors $X_j^{(i)} = \left(y_j^{(i)},...,y_{j+L_i+1}^{(i)}\right)^T \in R^{L_i}$, where $i=1,...,M$, L is the window length, and the number of columns is given by $K_i = N_i - L_i + 1$ and N is the number of data in the time series.

Then, decomposition is performed in a singular value for $X_V X_V^T$. In it, the eigenvalues of $X_V X_V^T$ are denoted by $\lambda_{V_1},...,\lambda_{V_{M \times L}}$, in decreasing

order of magnitude $\left(\lambda_{V_1} \geq ... \geq \lambda_{V_{M \times L}} \geq 0\right)$, and the eigenvectors are denoted by $U_{V_1},...,U_{V_{M \times L}}$. These elements are important to define the elementary matrices E_{V_i} according to $E_{V_i} = \sqrt{\lambda_{V_i}} U_{V_i} V_{V_i}^T$ and considering $V_V^T = X_V^T U_V / \sqrt{\lambda_V}$.

After the elementary matrices are defined, the two steps of the reconstruction stage are necessary: grouping and diagonal average. In the grouping step, the objective is to separate the elementary matrices into two groups: signal and white noise. Based on the measurement of the weighted correlation, as obtained by dividing the norm of groups by the internal product of a pair of groups, the elementary matrices are separated based on the lowest correlation obtained.

As each group is disposed in a multidimensional array composed of K_Y columns, the diagonal average is applied to convert the multidimensional array in a path vector that represents the original time series without white noise. According to the ANN-MSSA methodology, this series without white noise is used by the ANN prediction model.

Evaluation of the predictive performance

In this study, forecasts for ANN-MSSA and ANN are compared with those of the first 12 weeks subsequent to the final week of the sample. To evaluate the performance of the predictive models, the study makes used the measure of the mean-square error (MSE), as defined by:

$$EQM = \frac{1}{h} \sum_{N+1}^{N+h} \left(Y_j - \hat{Y}_j\right)^2 \qquad (5)$$

with Y_j representing the value of the original series, $\hat{Y}_j$ the value of the forecasting and h the amount of observations expected and selected for evaluation.

Application of statistical tests

Before the ANN-MSSA and ANN models were used, statistical tests were applied to know the time series characteristics. Thus, results obtained for the Anderson-Darling (AD) and Shapiro-Wilk (SW) tests (10% significance level) indicated that the time series were normally distributed. Considering the same significance level for the TSAY (1986) and MCLEOD & LI (1983) non-linearity tests, the series cannot be considered linear, which thus justify the use of the ANN model in the study. Finally, the Augmented Dickey–Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests to assess stationarity of

the time series confirm that they are not stationary. In addition to the tests, figure 1 shows that in the period Jan 13, 2012 to Dec 20, 2013, they are non-stationary time series due to the existence of both trends and some peaks in the time series.

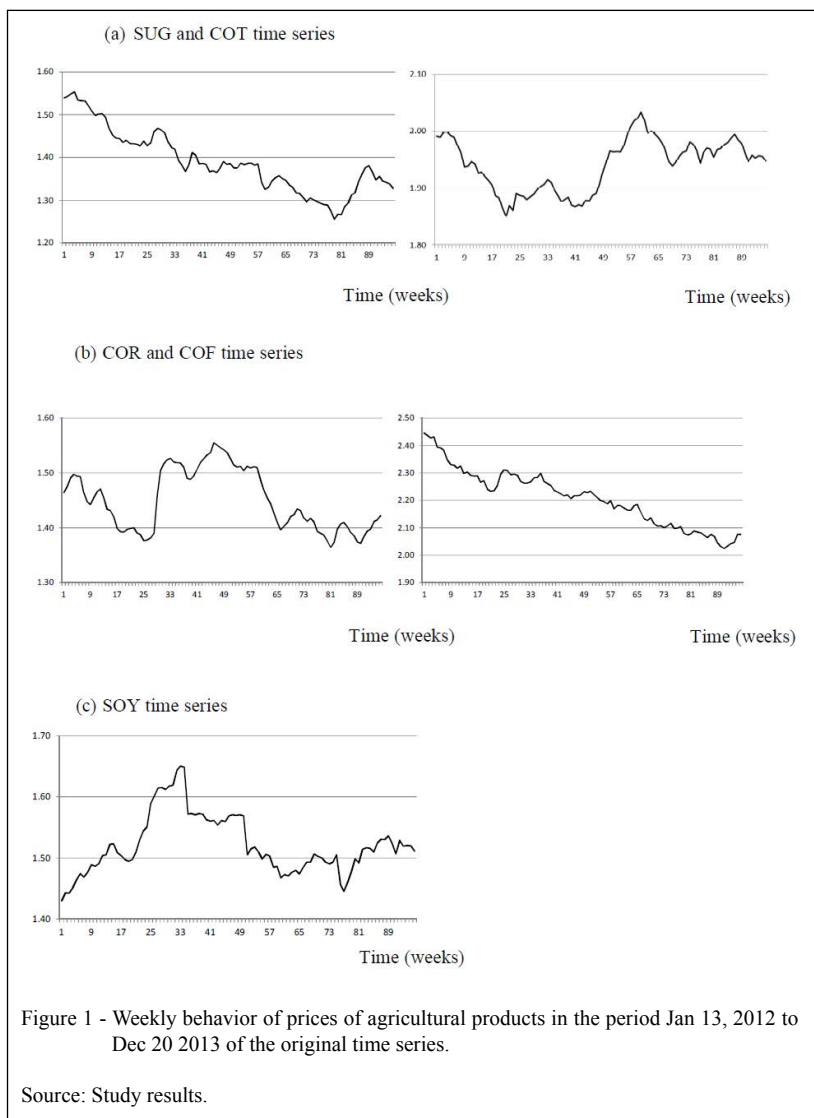Use of the ANN-MSSA methodology and ANN model

Based on the study by HASSANI & MAHMOUDVAND (2013), which indicated the ideal window size $L = \frac{N+1}{M+1}$ for MSSA, a value of 16 was used for this study. Then, 51 elementary matrices were obtained through the singular-value decomposition step. Reconstruction of the time series was the next step. Then, based on the weighted correlation measure, the 51 elementary matrices were separated, assigning values to SUG (14), COT (13), COR (12), COF (12), and SOY (15) to define the separation between signal and white noise.

This means that, in order to promote grouping in a time series, e.g., for SUG, the first 14 elementary matrices of 51 elementary matrices obtained in the singular-value decomposition step, correspond to the signal whereas the others (15-51) correspond to the white noise. After applying the diagonal average to the sum of these matrices, the time series without white noise is given by a column vector. Application of the MSSA model to ANN-MSSA is due to the fact that it favors capturing the structure of the study time series based on the structure of other time series.

After the white noise was eliminated, time series for each agricultural product was used in the ANN model using the R statistical program and ARNN package. This package uses a layer composed of some neurons, assuming a maximum a number of iterations equal to 100,000, with supervised training and use of the backpropagation algorithm. The first 80 of the 96 values available for each time series were used as an input and the 16 values remaining were used as a neural network training.

As the package allows changing the number of neurons, models with 5, 10, 15, 20, and 25 neurons were used Number of neurons was changed during the training phase, as in the studies of SOBREIRO et al. (2008), BRESSAN (2004), and FERREIRA et al. (2011).

After training, the model with 15 neurons showed the lowest value for the Akaike Information Criteria (AIC), being selected as the most appropriate for all the time series. In this study the difference between ANN-MSSA and the ANN model is given by

Figure 1 - Weekly behavior of prices of agricultural products in the period Jan 13, 2012 to Dec 20 2013 of the original time series.

Source: Study results.

the treatment to separate the noise from the original time series, i.e., for the ANN model, the time series were used without separating the white noise and multivariate analysis of the set of series.

RESULTS AND DISCUSSION

The forecasts obtained for MSSA-ANN using the error measure (5) correspond to the best performance (Table 1: BP) as compared to those obtained by the ANN model, because of the lowest values for error measurement.

The test proposed by DIEBOLD & MARIANO (1995) was applied to evaluate the statistical significance of the difference between the MSEs for the best-performance model and the

subsequent best-performance model. The results indicated that the null hypothesis for the models under comparison, by which the difference between error measurements is zero, could be rejected at 10%-significance statistical level for the SUG, COT, CORN, and SOY time series. The exception to this is given to the COF time series, whose best forecasting performance is given by the ANN model.

Besides the statistical test, which confirms the predictive superiority for the ANN-MSSA methodology. Figure 2 allowed to verify that the ANN-MSSA methodology was able to detect the trend prices for SUG, COT, CORN, and SOY.

Forecasting superiority of the ANN model can be seen in SOBREIRO et al. (2008).

Table 1 - Performance of price forecasting by the mean square error (MSE).

Series	L	h	ANN	-----------------------------MSE-----------------------------	
				MSSA-ANN	(BP)
SUG	16	12	9.8E-03	9.2E-04	MSSA-ANN
COT	16	12	3.0E-02	5.2E-03	MSSA-ANN
COR	16	12	7.6E-03	8.2E-04	MSSA-ANN
COF	16	12	6.6E-02	1.3E-02	SARIMA
SOY	16	12	1.6E-02	2.0E-04	MSSA-ANN

Source: Study results.

These authors confirmed the forecasting effectiveness for monthly sugar prices with the highest predictive approximation in recent months. In forecasting the weekly coffee and soybean prices, BRESSAN (2004) presented evidence favorable to the use of the classic (autoregressive) and ANN models as decision tools in agricultural negotiations.

The forecasting performance favorable to the ANN model is seen again FERREIRA et al. (2011). For this purpose, the authors used the corn monthly prices and concluded that the model showed a forecasting capacity, and thus it may be used in the decision-making process concerning the behavior of prices for agricultural products.

Prior to these studies, YONENAGA & FIGUEREDO (1998) showed that the ANN model is an alternative to the statistical methods to solve the problem of forecasting time series of agricultural products, and presents favorable predictive results in addition to properly indicat the trend in the real price. Other studies (LIMA et al. (2010); RIBEIRO et al. (2010); MIRANDA et al. (2013)) based on data from the national and international agricultural market (SHAHWAN & ODENING, 2007; LI et al., 2010; JHA; SINHA, 2013), indicated the ANN model as an alternative in forecasting agricultural prices.

In the international market, the prices for agricultural commodities were forecasted by a methodology that uses the classical autoregressive model combined with the ANN one, which was defined as a hybrid model in the study by SHAHWAN & ODENING (2007). For these authors, results for the hybrid model showed a better forecasting performance as compared to the autoregressive model. LI et al. (2010) were also based on daily and weekly prices for agricultural products in the international market, and concluded that the ANN model has a better predictive performance as compared to the autoregressive
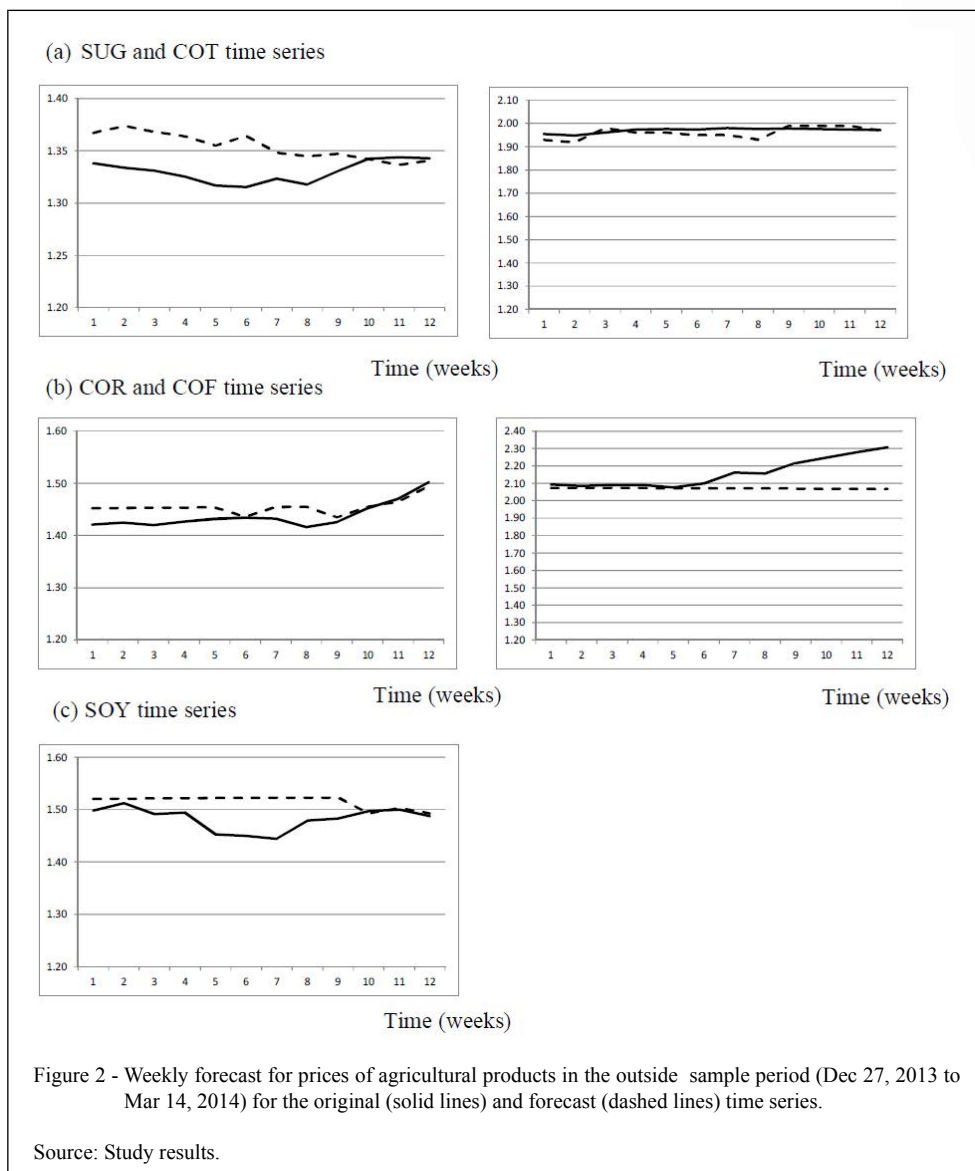
model in the short term. Justifying the importance of price management for farmers in the Indian market, the same conclusion was obtained by JHA & SINHA (2013) in their study on soybean and mustard prices.

Differently from the present study, the one mentioned above did not consider multivariate aspects, namely, the relationship between commodity prices. Nevertheless, their results do not reject using the ANN model, but improve its use when decomposition of the time series is included because the investigations on the national and international agricultural market do not take into account the set of prices for agricultural products. Thus, due to the multivariate nature of the MSSA model, which eliminates white noise and takes into account the relationship between prices for agricultural products, use of the MSSA-ANN may be considered an alternative that contributes to forecast the prices for agricultural products.

CONCLUSION

In the context of this study the empirical data showed that performance of the MSSA-ANN methodology is better than that of the ANN model, as the methodology allows obtaining a greater number of better performances for forecasts. Results obtained in the outside sample period, by using the MSE measurement and statistical test, confirm this conclusion.

This study combined the MSSA model to decompose the time series with the ANN model and added favorable evidence, showing to be another alternative in forecasting prices for agricultural products. Therefore, our results will be useful to formulate and implement policies directed to the agricultural sector. Thus, price forecasting obtained by the ANN-MSSA methodology is an alternative approach in financial planning, as it allowed to detect and forecast prices of the study agricultural products.

Figure 2 - Weekly forecast for prices of agricultural products in the outside sample period (Dec 27, 2013 to Mar 14, 2014) for the original (solid lines) and forecast (dashed lines) time series.

Source: Study results.

REFERENCES

BECKMANN, J.; CZUDAJ. R. Non-linearities in the relationship of agricultural futures prices. **European Review of Agricultural Economics**, v.41, n.1, p.1-23, 2014. Available from: <http://erae. oxfordjournals.org/content/41/1/1.full>. Accessed: Feb. 21, 2015. doi: 10.1093/erae/jbt015??.

BRESSAN, A.A. Tomada de decisão em futuros agropecuários com modelos de previsão de séries temporais. **Revista de Administração Eletrônica**, v.3, n.1, p.1-20, 2004. Available from: <http://www.scielo.br/scielo.php?script=sci_pdf&pid=S1676-64482004000100005&lng=en&nrm=iso&tlng=pt>. Accessed: Feb. 21, 2015. doi: 10.1590/S1676-56482004000100005.

CAMPOS, K.C. Análise da volatilidade de preços de produtos agropecuários no Brasil. **Revista de Economia e Agronegócio**, v.5, n.3, p.303-328, 2007. Available from: <http://ageconsearch.umn.edu/bitstream/54589/2/1_artigo. pdf>. Accessed: Feb. 21, 2015.

CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA (CEPEA). Available from: <http://www.cepea.esalq. usp.br/cepea>. Accessed: Feb. 21, 2015.

CERETTA, P.S. et al. Previsão do preço da soja: uma comparação entre os modelos ARIMA e redes neurais artificiais. **Informações Econômicas**, v.40, n.9, p.15-27, 2010. Available from: <ftp://ftp. sp.gov.br/ftpiea/publicacoes/ie/2010/tec2-0910.pdf>. Accessed: Feb. 21, 2015.

DIEBOLD, F., MARIANO, R. Comparing predictive accuracy. **Journal of Business and Economic Statistics**, v.13, n.3, p.253-265, 1995. Available from: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.352.9389&rep=rep1&type=pdf>. Accessed: Feb. 21, 2015. doi: 10.1.1.352.9389.

FERREIRA, L. et al. Utilização de redes neurais artificiais como estratégia de previsão de preços no contexto de agronegócio. **Revista de Administração e Inovação**, v.8, n.4, p.6-26, 2011. Available from: <http://www.revistas.usp.br/rai/article/view/79235/83307>. Accessed: Feb. 21, 2015. doi: 10.5773/rai.v8i4.475.

HASSANI, H.; MAHMOUDVAND, R. Mutivariate singular apectrum analysis: a general view and new vector forecasting approach. **International Journal of Energy and Statistics**, v.1, n.1 p.55-83, 2013. Available from: <http://www.researchgate.net/profile/Hossein_Hassani7/publication/263870252_MULTIVARIATE_SINGULAR_SPECTRUM_ANALYSIS_A_GENERAL_VIEW_AND_NEW_VECTOR_FORECASTING_APPROACH/links/53ec91630cf2233164944e12.pdf>. Accessed: Feb. 21, 2015. doi: 10.1142/S2335680413500051.

HAYKIN, S. **Redes neurais**: princípios e prática. 2.ed. Porto Alegre: Bookman, 2001. 900 p.

JHA, G.K.; SINHA, K. Agricultural price forecasting using neural network model: An innovative information delivery system. **Agricultural Economics Research Review**, v. 26, n.2, p.229-239, 2013. Available from: <http://ageconsearch.umn.edu/bitstream/162150/2/8-GK-Jha.pdf>. Accessed: Feb. 21, 2015.

LI,G-Q.etal.Short-termpriceforecastingforagro-productsusingartificial neural networks. **Agriculture and Agricultural Science Procedia**, v.1, n.2010, p.278-287, 2010. Available from: <http://ac.els-cdn.com/S2210784310000367/1-s2.0-S2210784310000367-main.pdf?_tid=d729c7b8-7f62-11e6-9f6e-00000aab0f26&acdnat=1474397497_c8c47e606498f516426f5168dd65e40b>. Accessed: Feb. 21, 2015. doi:10.1016/j.aaspro.2010.09.035.

LIMA, F.G. et al. Previsão de preços de commodities com modelos ARIMA-GARCH e redes neurais com ondaletas: velhas tecnologias – novos resultados. **Revista de Administração**, v.45, n.2, p.188-202, 2010. Available from: <http://www.revistas.usp.br/rausp/article/view/44503/48123>. Accessed: Feb. 21, 2015. doi: 10.1590/S0080-21072010000200008.

MARTINS, T.M.; MARTINELLI, D.P. Ciclos e previsão cíclica dos preços de commodities: um modelo indicador antecedente para commodity açúcar. **Revista de Administração, Contabilidade e Economia**, v.1, n.2, p.2-12, 2010. Available from: <http://www.fundace.org.br/artigos_racef/artigo01_02_2010.pdf>. Accessed: Feb. 21, 2015.

MCLEOD, A.I.; LI, W.K. Diagnostic checking ARMA time series models using squared residual autocorrelations. **Journal of Time Series Analysis**, v.4, n.4, p.169-176, 1983.

MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO (MAPA). Available from: <http://www.agricultura.gov.br/>. Accessed: Feb. 21, 2015.

PASQUOTTO, J.L.D. **Previsão de séries temporais no varejo brasileiro**: uma investigação comparativa da aplicação de redes neurais recorrentes de Elman. 2010. 191f. Dissertação (Mestrado em Administração) - Faculdade de Economia, Administração e Contabilidade de São Paulo, USP, São Paulo, SP.

PINHEIRO, C.A.O.; SENNA, V. Previsão de preços para commodities agrícolas através de redes neurais e análise singular multivariada. In: ENCONTRO DA ASSOCIAÇÃO NACIONAL DOS PROGRAMAS DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO, 39., 2015, Belo Horizonte, MG. **Anais…** Belo Horizonte: ENANPAD, 2015. p. 1-13.

RIBEIRO, C.O. et al. Um modelo hierárquico para previsão de preços de commodities agrícolas. **Revista Produção On-line**, v.10, n.4, p.719-733, 2010. Available from: <http://producaoonline.org.br/rpo/article/view/225/734>. Accessed: Feb. 21, 2015. doi: 10.14488/1676-1901.v10i4.225.

SHAHWAN, T.; ODENING, M. Forecasting agricultural commodity prices using hybrid neural networks. In: CHEN, S.H. et al. **Computational intelligence in economics and finance**. Berlin: Springer, 2007. v.2, p.63-74.

SOBREIRO, V.A. et al. Uma estimação do valor da commodity de açúcar usando redes neurais artificiais. **Revista P&D em Engenharia de Produção**, v.6, n.1, p.36-53, 2008. Available from: <http://www.revista-ped.unifei.edu.br/documentos/V06N01/n07_art03.pdf>. Accessed: Feb. 21, 2015.

TSAY, R. Non-linearity tests for time series. **Biometrika**, v.73, n.2, p.461-466, 1986.

YONENAGA, W.H.; FIGUEIREDO, R.S. Previsão do preço da soja utilizando redes neurais. In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 19., 1999, Rio de Janeiro, RJ. **Anais...** Rio de Janeiro: ENEP, 1999. p. 1-17.